DECISION TREE CLASSIFICATION OF MULTIDATE AVIRIS DATA FOR MAPPING WOODLAND ENCROACHMENT INTO THE GREAT PLAINS

David Meyer, Bruce Wylie, Cullen Robbins[1]
Lee Vierling, Patrick Kozak[2]

1. BACKGROUND

The Great Plains of North American has experienced a significant increase in woodland cover over the last 200 years. This encroachment of woodlands into grasslands, as well as the "thicketization" of savannas has been observed worldwide (Archer et al., 1995). Several reasons have been offered to explain this transformation. The traditional view holds that changes in land management are the cause, operating through several modes of disturbance. A primary mode is a reduction in burn frequency, coupled with increased grazing and tree planting. Burning tends to favor native prairie species (suppressing encroachment by woodland species), while grazing tends to select for the unpalatable woodland species (Reich et al., 2000). Under these conditions, species introduced as shelterbelts and windrows can spread into prairie ecosystems. A secondary mode of disturbance is the abandonment of agricultural land, which, coupled with fire suppression, favors the development of woodland cover. Climatic factors have also been proposed. For example, the replacement of C_4 grasses by C_3 shrubs in North America inferred over the last 200 years could either be due to systems that were established at the end of the "little ice age" and only marginally supported climatically at the time of settlement (Neilson, 1986), or due to post-industrial CO_2 fertilization (Idso, 1992). The latter proposition is difficult to support, however. Even though the global nature of encroachment supports the fertilization hypothesis, a number of alternative explanations are available (), undermining this point of view.

The implication of these changes is still poorly understood. A simple view holds that the transformation of grassland to woodland increases carbon sequestration. Indeed, the expansion and densification of woodlands has been proposed as an explanation for up to 50% of the total US carbon sink associated with land cover changes (Houghton et al., 1999). This transformation has been implicated in changes in meteorological and climatic patterns within the Great Plains, resulting from alterations to water and energy fluxes between the surface and the atmosphere, as well disturbances to the carbon and nitrogen exchange cycles as grasslands give way to wooded cover. Although carbon sequestration is an important tool for reducing greenhouse gases, this sequestration occurs at the expense of soil carbon (Sage et al., 1999) as grasslands are transformed. In areas prone to droughts and frequent fires, the resilience of carbon stored in tree biomass (wood) as compared to soil carbon is questioned.

In an attempt to monitor this encroachment, it is important to map the distribution of species that may indicate the early onset of grassland-woodland transformation. One such species is the eastern red cedar (*Juniperus virginiana*). Extensive throughout the eastern and central US, this juniper species has spread extensively into the central Great Plains. Because it's drought resistance, highly adaptive physiology and extensive root system, the eastern red cedar has been widely used to control soil erosion and to reduce the desiccating effects of wind. Introduced as windrows and shelterbelts, the species will usually give way to climax deciduous and conifer species. Although considered shade intolerant, eastern red cedar can still flourish under lower- to mid-density deciduous stands, due to it's low capacity for water loss and ability to photosynthesize adequately during leaf-off overstory conditions. The species is highly sensitive to fire, it's primary control, but as fire is suppressed the juniper can spread from shelterbelts into the surrounding pastureland and abandoned cropland. Once established in these new areas, they are usually replaced by more fire-tolerant hardwoods and pines, although they can occasionally outlive competing hardwoods through allelopathy and a highly competitive physiology in marginal soil and climatic conditions.

---

[1] Raytheon Corp. USGS EROS Data Center (dmeyer@usgs.gov)
[2] South Dakota School of Mines and Technology

Mapping the eastern red cedar throughout the Great Plains using remote observation techniques can be problematic for several reasons. First, in the early stages of encroachment, the species is either hidden under deciduous overstory, or expressing itself only through isolated, patchy occurrences in open areas. Its spectral signature is not always distinct from other conifer species as viewed by lower-dimensional multispectral instruments, although throughout much of the plains, there are few other conifers co-occurring with the juniper in large numbers. However, one of its prime expansion corridors is along the riparian systems that cut west-to-east across the central plains. At some point, the eastern red cedar can co-occur with western conifers, such as the ponderosa pine. This is the case along the Niobrara River in Nebraska, a unique riparian system where eastern hardwoods can intermix with western conifers. The Niobrara and other rivers are important conduits to juniper expansion to the west, but present all the problems associated with remote detection: patchiness, overstory obscuration and spectral confusion. The problem of remote detection is approached here using high spectral resolution to improve spectral discrimination, and leaf-on, leaf-off data acquisitions at high spatial resolution to identify the eastern red cedar under varying conditions of background spectra and overstory obscuration. The primary instrument for this study is the Advanced Visible and Infrared Imaging Spectrometer (AVIRIS) coupled with field observations of juniper stands.

2. AVIRIS DATA COLLECTION AND PREPARATION

The geographic focus of this study is the Niobrara Valley Nature Preserve in north central Nebraska, owned and managed by The Nature Conservancy. The preserve includes the intermixed riparian woodland described above, as well as extensive tracts of the Nebraska Sand Hills to the south of the river. Two low-altitude (15,500 ft above mean terrain) Twin Otter AVIRIS acquisitions occurred in 1999: July 22 and November 11. The first acquisition captures the grassland and woodlands under conditions of maximum green cover (the native prairie grasses are dominant warm-season), while the latter captures senesced grasses and leafless hardwood stands. Over 600 field control points were measured over the preserve to include a variety of cover types: grasslands, woodland, shrub (primary sumac), and bare soil. The woodland class was further subdivided into deciduous, conifer-ponderosa pine, conifer-juniper and mixed classes. These field points were located to an accuracy of around 2 meters and used to train the classification described below.

Although the AVIRIS data were geometrically corrected before delivery, the inherent errors were sufficiently large to require further correction. Both the scenes were re-transformed using a standard US Geological Survey digital orthophoto quadrangle (DOQ) as control. Although the scene-to-scene accuracies were high after correction, the DOQ's intrinsic locational uncertainty of around 6 meters must be considered in light of the AVIRIS image's 2.7 meter resolution and the locational accuracies of the field control used in the classification. These problems come into play during the classification accuracy assessment described later. Using field spectroscopy measurements, the radiance-calibrated imagery were corrected for the atmosphere and converted to (isotropic) surface reflectance. Minimum noise fraction (MNF) transformations were applied to both data sets, and an MNF threshold of 2.0 was used to retain MNF components for classification.

3. CLASSIFICATION METHODOLOGY

Two supervised classification methods were used, one using matched filtering for each cover type under investigation, and the other using decision trees (Friedl et al., 1999). For the matched filtering approach, regions of interest (ROIs) were extracted from the imagery using the global positioning system (GPS) locations acquired during the field survey. Image spectra from these ROI's were used as reference spectra for the matched filter algorithm, and a matched filter image was computed for each of the cover types under study. The performance of this approach was used by observing the value of the matched filter values for the different classes at control GPS field survey locations withheld from the training set locations for the purpose of validation. The means by which the performance of this method is measured is described in the next section. The matched filtering approach for the November dataset was not complete at this writing and is not included here.

The decision tree approach used the C5.0 univariate decision tree algorithm. Decision tree classifiers have become increasingly popular for several reasons. First, they are non-parametric, requiring no prior assumptions regarding probability density functions of a given data ensemble. Also, they are considered to be relatively robust with respect to nonlinear and noisy interrelationships between features expressed in a given data set (Friedl et al., 1999). The decision tree algorithm works by computing a metric known as the information gain ratio, a measure of the reduction in entropy produced by subdividing ("splitting") data into subsets based on feature vector decision thresholds, based on a given set of training signatures. The "tree" nomenclature describes the morphology of the splitting process: recursive binary decisions resulting in a number of subdivision. The final tree configuration is achieved by maximizing the information gain threshold at each node (binary decision) in the tree. Performance of the classification is based on standard classification validation procedures.

The July 22 classification was described previously and is synopsized in the next section. For the classification of the November 11 data set, two classification schemas were used. The first was for a general classification scheme dividing the image into 5 basic types (shadow, bare, water, woody, and grass), and the second was a woody-specific division to determine which types of woody species were present (cedar, pine, deciduous, shrub, and mixed (dec. and cedar)). The desired separations were run through the C5.0 decision tree to determine the accuracy and usefulness of the separation. The following separations were run: general classification, cedar/other/mixed, woody/other separation. Cross-validation was used to estimate error rates of various landcovers with different decision tree parameters. Boosted decision trees () produced the most useful results. When the results were determined to be meaningful, the decision tree was run on the entire 32-band MNF image, since it had been trained by the points. This produced classified images and confidence maps.

Due to lower accuracy than expected in some of the images, it was decided to aid improve classification using results from the July classification. Only the woody class was used from the July result as a pre-stratification for the November classification. The mixed class from the July result was excluded to reduce training errors. The training points were again used to estimate the ability of decision tree to classify the image using cross-validation. The decision tree was then run spatially on the entire 32-band November MNF image. The corresponding woody points for both the leaf on and leaf off images were matched up and run through decision tree in an attempt to improve our woody separation. This decision tree was also run spatially on both images and a composite was produced.

A final classification was then produced using a decision model combining the aforementioned results. The general classification was used for the classes shadow, bare, water, and grass. If the leaf-on result called a pixel deciduous and the leaf-off called it cedar or mixed, it was put in a mixed class. If a pixel was deciduous in the leaf-on and pine in the leaf-off, it was also added to the mixed class. Of the remaining pixels, if it was deciduous in the leaf on, it was given a deciduous class. If it was cedar in leaf-off, it was classified as cedar. Likewise, if it was pine or shrub in leaf off, these classes were created and named correspondingly. If a pixel was classified as mixed in the leaf off, it was added to the mixed class. The remaining pixels were classified according to how they were classified in the on and off composite image. (Fig. 1)

4. RESULTS

The results from the July 22 classification, reported previously reported that both the matched filter and decision tree approaches were able to discriminate between cedar from other cover types with accuracies of about 85%.

The cross-validation step in the decision tree provides statistics in the form of a confusion matrix with which an accuracy assessment can be done. The matrices from all of the different decision tree outputs were combined using area weighting to produce an accuracy assessment for each class in the final image. The overall accuracy of the classification is 84 %. The shadow and water classes had the highest accuracies, 94 % and 93 %, respectively. The bare and grass classes had accuracies of 83% and 89%, respectively. Cedar was accurately classified 52% of the time, while the mixed class was accurately classified 35% of the time. Coniferous trees were accurately classified 46% of the time. Deciduous trees and shrubs were accurately classified 64% and 62% of the time, respectively. The following chart shows the breakdown of the above:

Table 1: decision tree results

class	% accuracy
Shadow	94
Water	93
Bare	83
Grass	89
Cedar	52
Coniferous	46
Deciduous	64
Shrub	62
Mixed	35
Overall	84

5. DISCUSSION

The low cedar classification accuracy for the November 11 acquisition is the result of several complicating factors. A primary factor is the overall locational uncertainties in the surveyed control points and the imagery used for analysis (including the reference DOQ). Because of the patchiness of the exposed juniper stands and the 2.7 m resolution of the Twin Otter AVIRIS data, small locational errors in field surveys and imagery all could lead to large errors in a supervised, multi-date classification strategy. Another influence is the general illumination conditions of the November 11 overflight: low sun (shadowing, anisotropic reflectance effects). Also, attempts to measure through obscuring hardwood canopies may be somewhat ambitious, even under leaf-off conditions. Finally, the quality of the field observations must be considered: in a very patchy environment, what is considered "quality" control? Stands of pure juniper, large enough to be spectrally unambiguous, are extremely rare.

Methods for mitigating these factors include the use of precise control for image registration, more accurate field measurements and better examples of "target" spectral signatures for classification. Due to the reasons mentioned above, pure spectral signatures may be unobtainable from the image, so strategies employing the use of reference field spectra along with unmixing strategies should be pursued further.

6. REFERENCES

Archer, S., D.S. Schimel and E.A. Holland, 1995, "Mechanism of Shrubland Expansion: Land Use, Climate or CO_2?," Climate Change, vol. 29, p. 91-99.

Friedl, M.A., C.E. Brodley and A.H. Strahler, 1999, "Maximizing Land Cover Classification Accuracies Produced by Decision Trees at Coontinental to Global Scales," IEEE Trans. Geoscience & R.S. vol. 37, no. 2, pp. 969-982.

Houghton, R.A., J.L. Hackler, and K.T. Lawrence, 1999, "The US Carbon Budget: Contributions from Land-Use Change," Science, vol. 285, pp. 574-578.

Idso, S.B., 1992, "Shrubland Expansion in the American Southwest," Climate Change, vol. 22, p. 85-86.

Neilson, R.P., 1986, "High Resolution Climatic Analysis and Southwest Biogeography,", Science, vol 232, pp. 27-34.

Reich, P.B., D.W. Peterson, D.A. Wedin, and K. Wragge, 2000, "Productivity and Nitrogen Cycling across a Fire Dominated Forest-Grassland Continuum," Ecology, in press.

Sage. R.F., D.A. Wedin, and M. Li, 1999, "The Biogeography of C_4 Photosynthesis: Patterns and Controlling Factors," in R. F. Sage and R.K. Monson, The Biology of C_4 Photosynthesis, Academic Press, pp. 313-373.

7. FIGURES

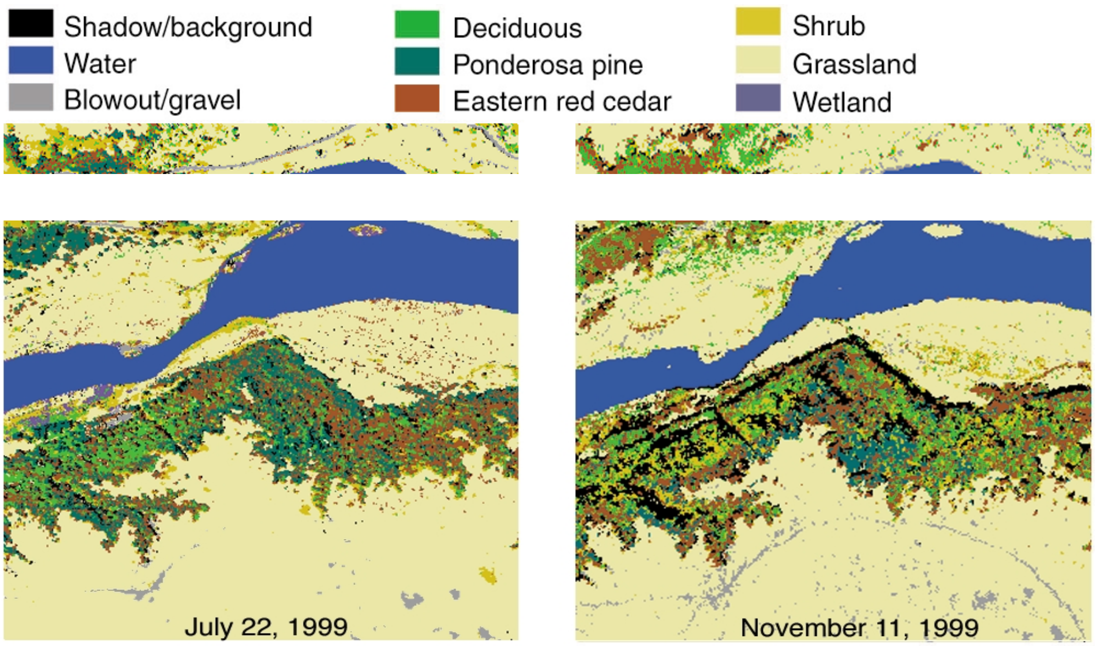

Figure 1.
Classification results from the univariate decision tree for both dates.